AN INTRODUCTION TO ECONOMETRICS

Oxbridge Economics; Mo Tanweer
Mohammed.Tanweer@cantab.net

Econometrics

- What is "econometrics"?
  - Econometrics means "economic measurement"
  - Economics + Statistics = Econometrics
- Econometrics is concerned with the tasks of developing and applying **quantitative or statistical** methods to the study of economic principles
- Econometrics is the use of statistical techniques to analyse economic data and compare with economic theory
- One of its aims is to give empirical content to economic theory
- What makes Econometrics different to Statistics?
  - Economic theories tend to be qualitative in nature
  - Economic data tends to be observational and more complex

Examples of econometrics

- As economists, we would like to understand the relationship between economic variables.
  - Is human capital a fundamental cause of growth?
  - Do improvements in educational spending by governments improve academic performance?
  - Can increases in the price of oil lead to reductions in national income?
  - Does an increase in national savings lead to an increase in investment?
  - If a government decreases the duration of time it offers unemployment benefit, does this lead to a lower unemployment rate?
  - Do fertility decisions in Pakistan exhibit socio-economic conforming behaviour?

Regression Analysis

- When we consider the nature and form of a relationship between any two or more variables, the analysis is referred to as regression analysis.

- Theoretical econometrics:
  - considers questions about the statistical properties of estimators and tests,

- Applied econometrics:
  - is concerned with the application of econometric methods to assess economic theories
    - Descriptive - How does the stock market and interest rate move together?
    - Forecasting - Will we have a recession next year?
    - Causal - If we raise the minimum wage, would unemployment soar?

Data types

E.g. Incomes for a country

- **Time-series**

  - A data set containing observations on a single phenomenon observed over multiple time periods is called time series

  - In time series data, both the values and the ordering of the data points have meaning

- **Cross-section**

  - A data set containing observations on multiple phenomena observed at a single point in time is called cross-sectional

  - In cross-sectional data sets, the values of the data points have meaning, but the ordering of the data points does not.

- **Panel / Longtitudinal / TSCS**

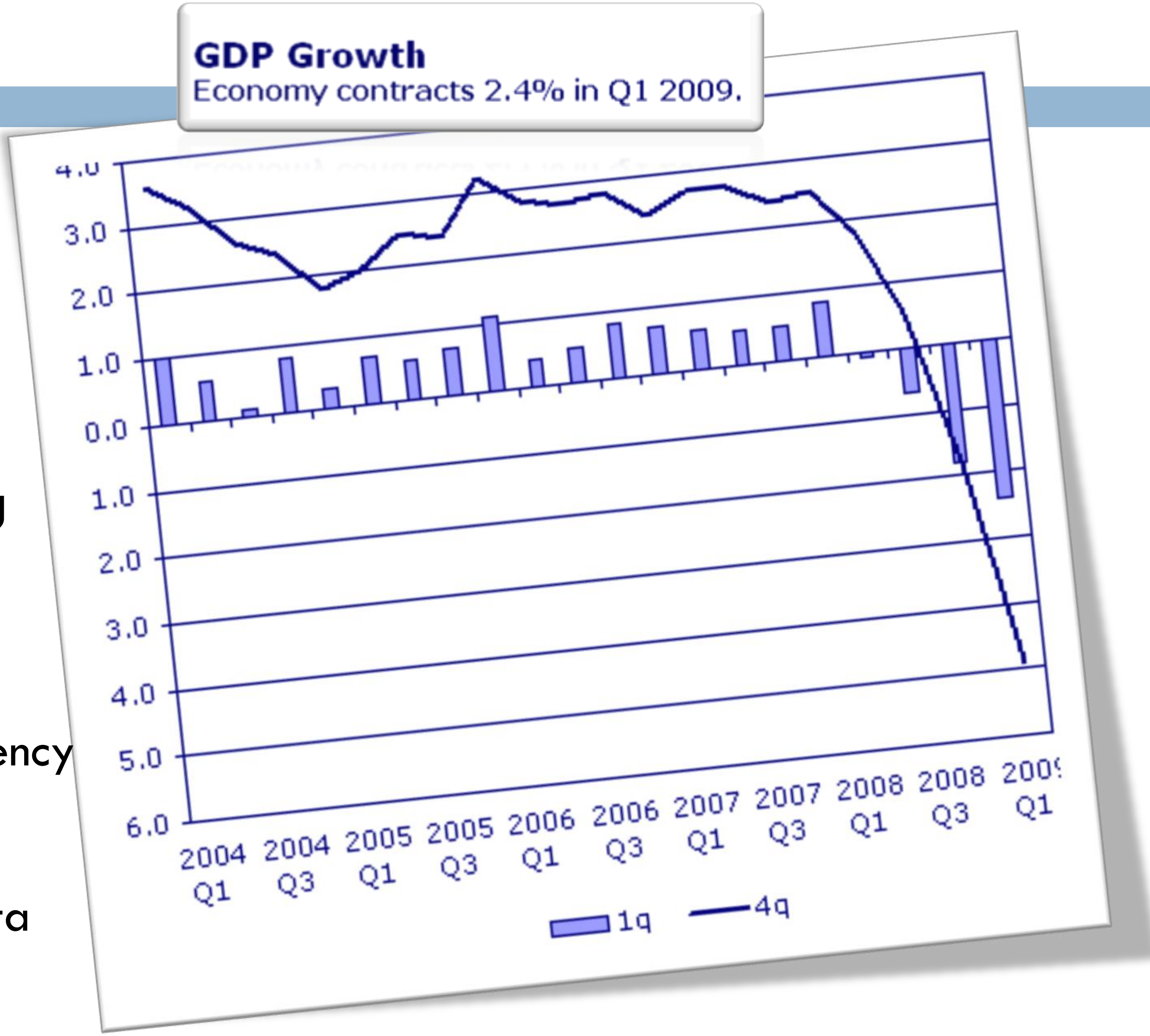
  - Two-dimensional data

  - A data set containing observations on multiple phenomena observed over multiple time periods is called panel data

Time-series

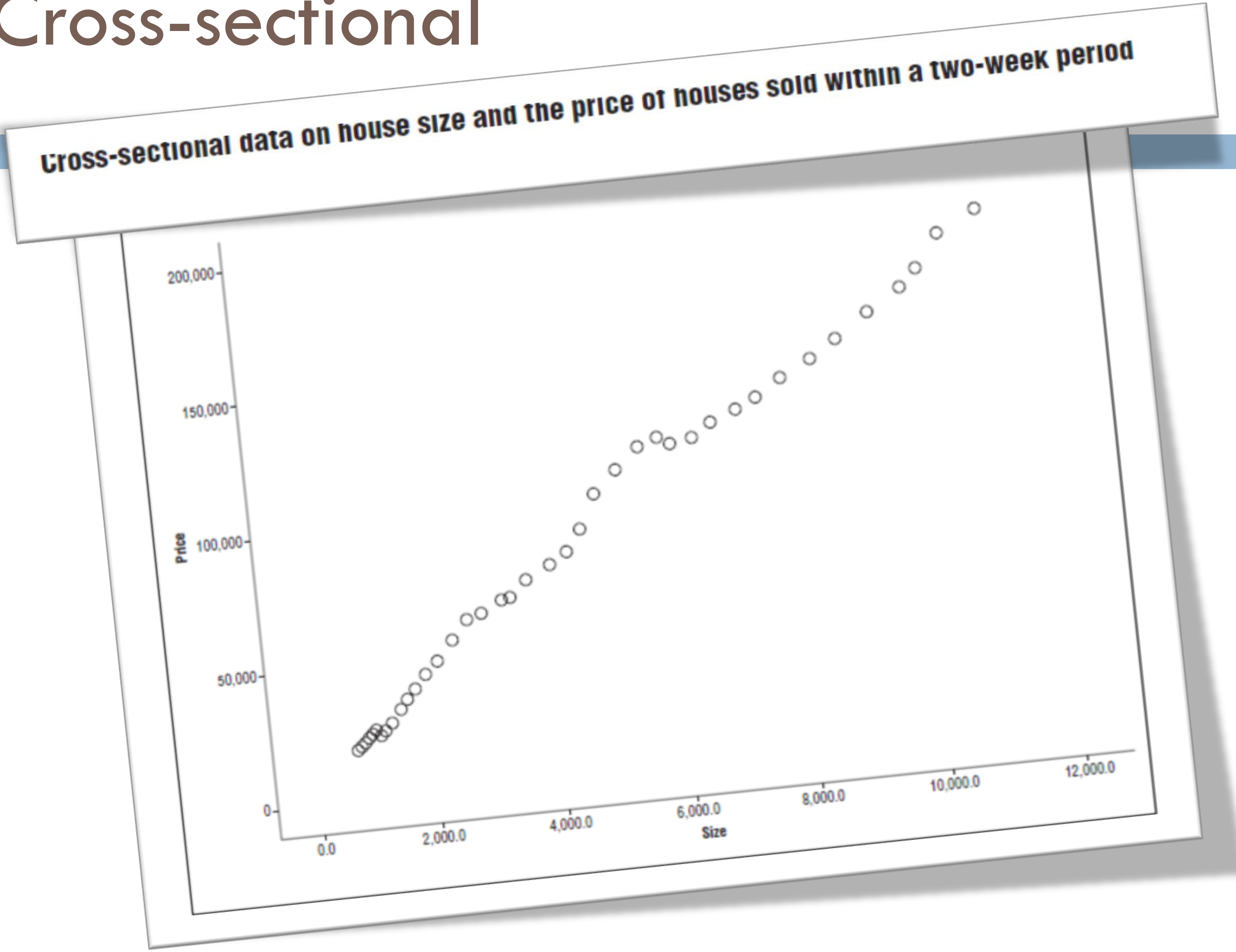
- Forecasting
- Chronology matters
- Interdependency issue
- Seasonal data

Time-series



- Forecasting
- Chronology matters
- Interdependency issue
- Seasonal data

Cross-sectional



Panel/Longitudinal

- The Panel Study of Income Dynamics

- The PSID had collected information on more than 70,000 individuals spanning as many as 4 decades of their lives.

- measures economic, social, and health factors over the life course and across generations

  - Do fertility decisions in Pakistan exhibit socio-economic conforming behaviour?

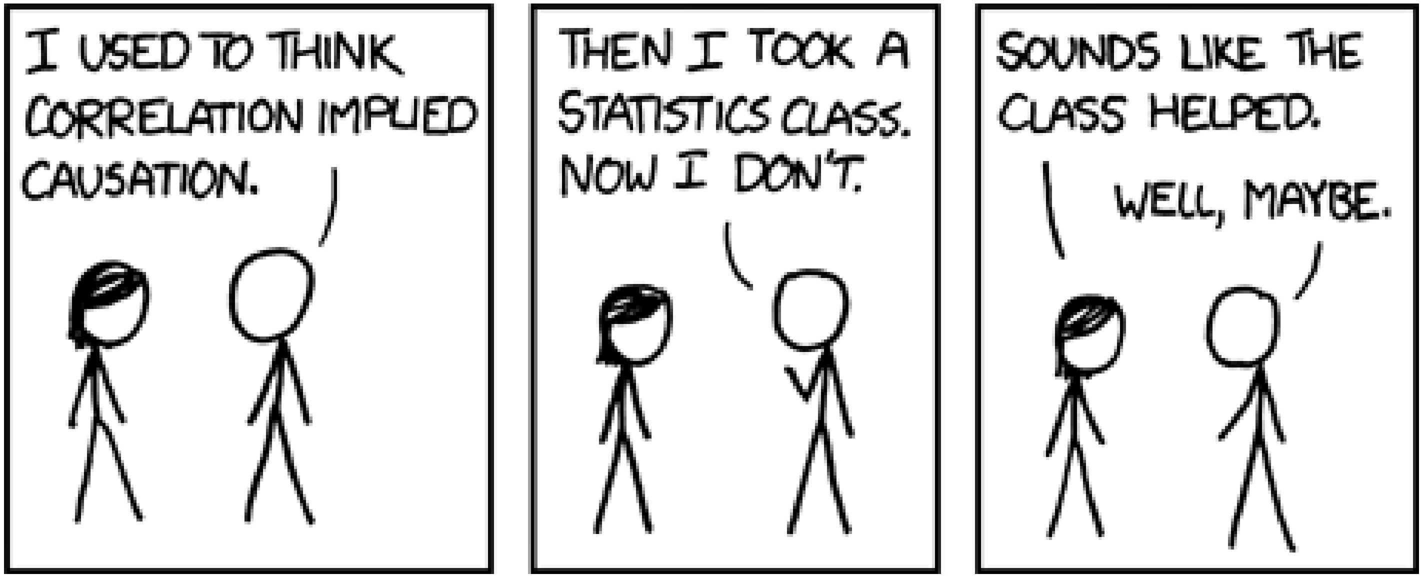
    - Write down a list of data variables you would like in your model

Statistical Packages

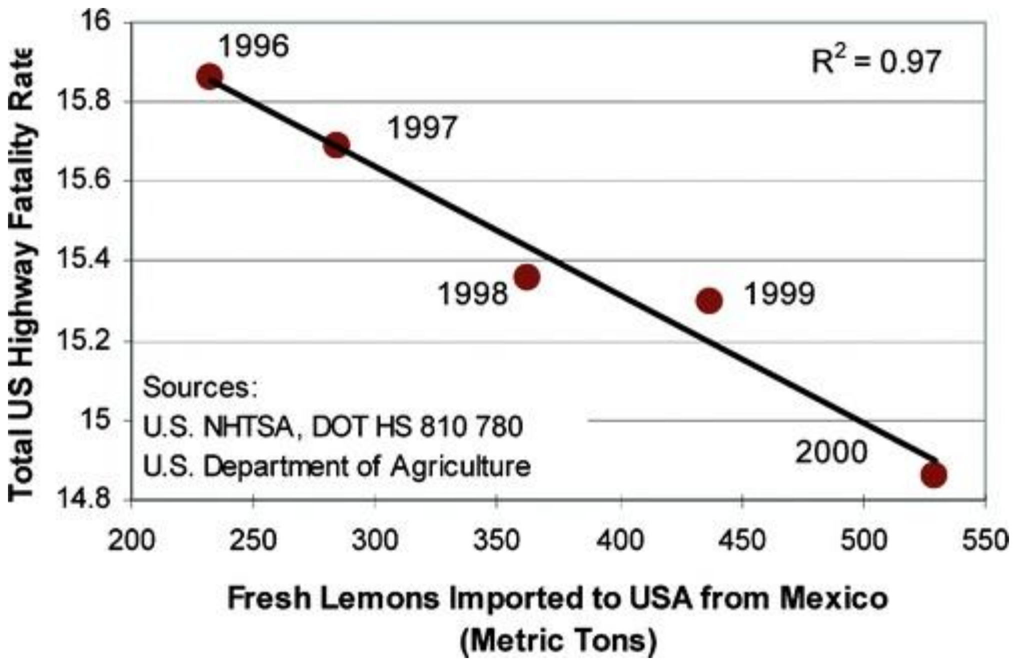
- EViews
- SPSS
- Microfit
- Stata
- PcGive
- Minitab
- Shazam
- Excel

Correlation vs Causation

- Correlation is not necessarily causation.
  - Post hoc ergo propter hoc / Cum hoc ergo propter hoc



Correlation vs Causation



- A occurs in correlation with B.
- Therefore, A causes B.
- This is a logical fallacy because there are at least four other possibilities:
  - B may be the cause of A
  - some unknown third factor C is actually the cause of both A and B
  - B may be the cause of A at the same time as A is the cause of B
  - coincidence

Be careful what you infer from your statistical analyses.
Be sure your relationship makes sense.

Econometrics in practice

- The relationship between Income and Consumption.

- Economic theory:
  - Keynes postulated a positive relationship between consumption and income
  - General Theory of Employment, Interest and Money.
  - ***"The fundamental psychological law… is that men [women] are disposed, as a rule and on average, to increase their consumption as their income increases, but not as much as the increase in their income"***

- Statistics :
  - Let's get the data and look at the causal relationship

Income and Consumption

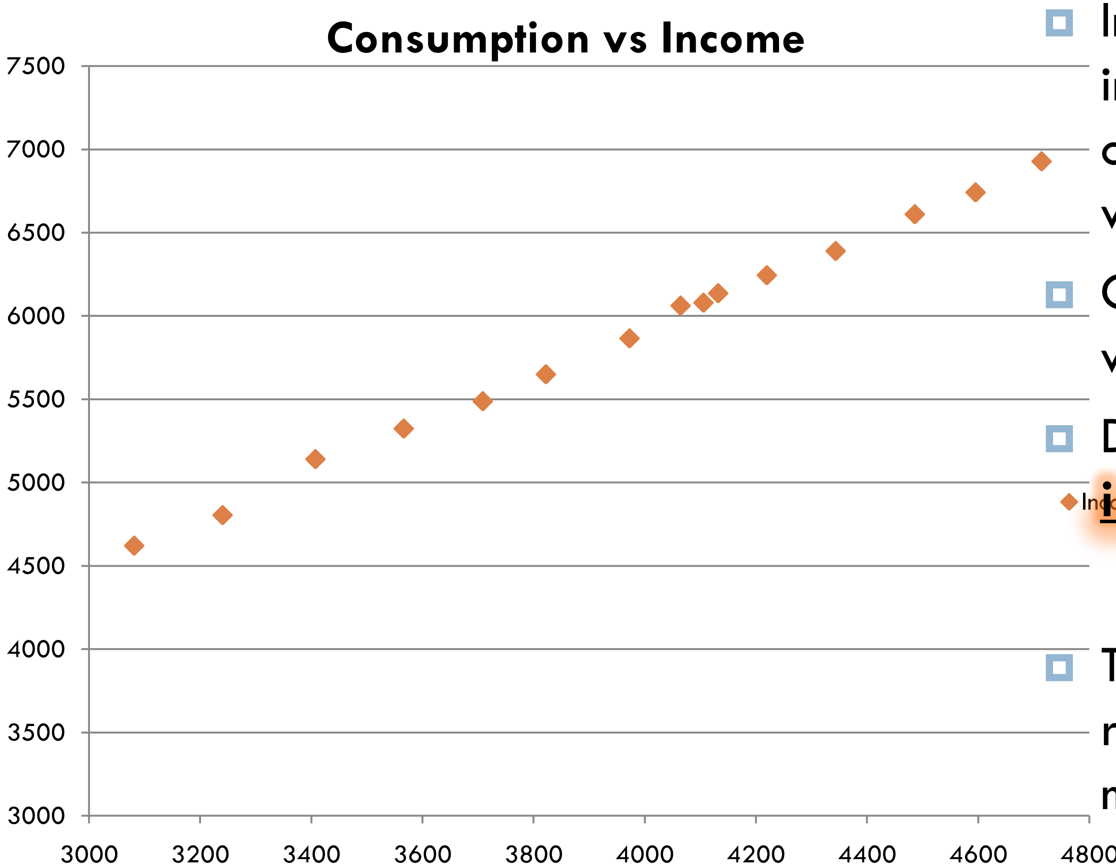
Data on Personal Consumption Expenditure
And Gross Domestic Product;1982-1996)
all in 1992 billions of dollars

	C	GDP
1982	3081.5	4620.3
1983	3240.6	4803.7
1984	3407.6	5140.1
1985	3566.5	5323.5
1986	3708.7	5487.7
1987	3822.3	5649.5
1988	3972.7	5865.2
1989	4064.6	6062
1990	4132.2	6136.3
1991	4105.8	6079.4
1992	4219.8	6244.4
1993	4343.6	6389.6
1994	4486	6610.7
1995	4595.3	6742.1
1996	4714.1	6928.4

- Casual observation suggests that the relationship between income and consumption is positive in that consumption rises as income rises.
- However, we want to analyse more formally if a relationship exists.
- We use regression analysis to look at this relationship formally.

Income and Consumption

- Consumption is the **dependent** variable)



- In regression analysis an important issue is the "direction of causation" between variables.
- Consumption is the **dependent** variable (Y)
- Disposable Income is the **independent** variable (X)
- This confirms the positive relationship but we need to be more **rigorous** in our analysis

- Disposable Income is the **independent** variable

The Theory: Keynes

- Let us assume that the relationship between consumption and income takes the form of the Keynesian consumption function:

$$Y = \alpha + \beta_1 X$$

  - Where β_1 is between 0-1

- α and β_1 are known as the PARAMETERS of the model (a.k.a the "intercept" and "slope coefficients")
- Y is the dependent variable, in this case, consumption
- X is the independent variable, in this case, income
- β_1 is a measure of what?

Terminology

- Dependent variable
  - Explained variable
  - Predictand
  - Regressand
  - Endogenous
  - Controlled variable
- Independent variable
  - Explanatory variable
  - Predictor
  - Regressor
  - Exogenous
  - Control variable

Econometrics

$$E(Y) = \alpha + \beta_1 X$$

is the POPULATION regression equation

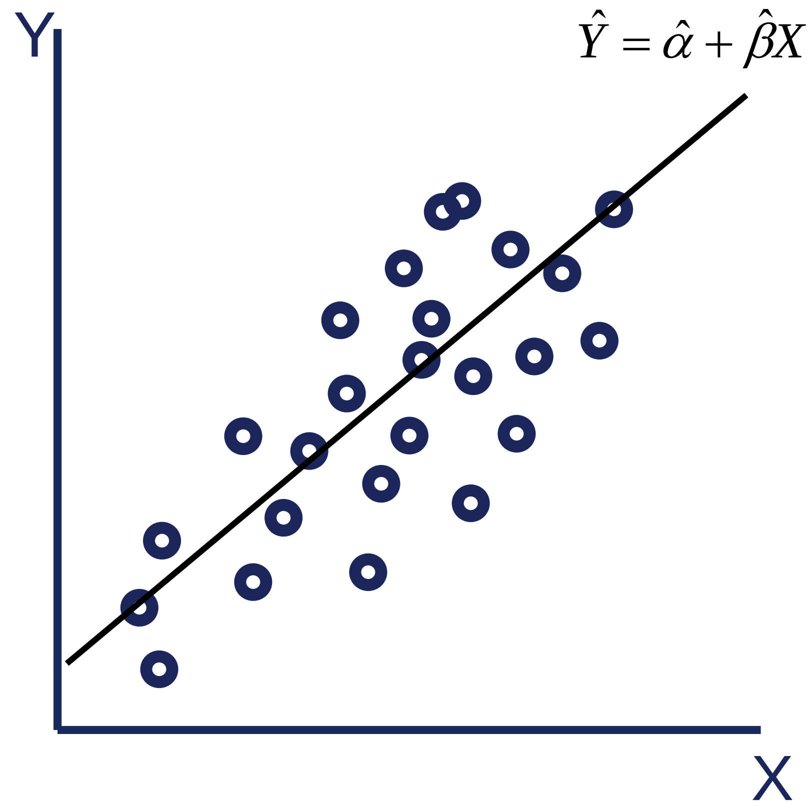
- The actual consumption Y of a household will not always equal its expected value E(Y).
- Actual consumption of a household may be 'disturbed' from its expected value by any one of innumerable factors, and we shall therefore write actual consumption as:

$$E(Y) = \alpha + \beta_1 X + u$$

- The disturbance (u) (or, e, error term) represents the effect on household consumption of all variables other than income.

Econometrics

- The population regression equation is unknown to any investigator, and remains unknown.
- Therefore have to fit a straight line to the scatter points. This line can then be regarded as an "estimate" of the population equation.



- The fitted line we write as:

$$\hat{Y} = \hat{\alpha} + \hat{\beta}X$$

- The sample regression equation represents a straight line with intercept "alpha-hat" and slope "beta-hat".
- "Y-hat" is known as the predicted value (or "estimate")of Y

Econometrics

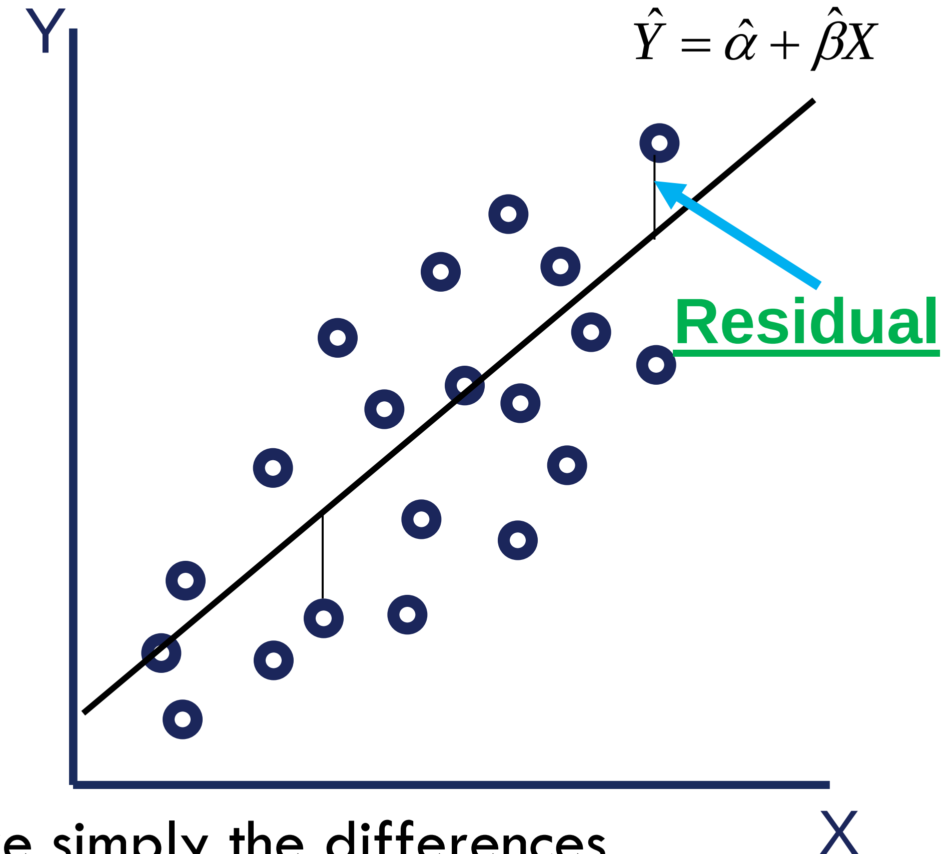
$$Y = \alpha + \beta_1 X$$

$$\hat{Y} = \hat{\alpha} + \hat{\beta} X + \hat{u}$$

$$Y = \hat{Y} + \hat{u}$$

$$\hat{u} = Y - \hat{Y}$$

$$\hat{u} = Y - \hat{\alpha} - \hat{B}_1 X$$



- So the residuals (or error terms) are simply the differences between the actual and estimated Y values.
- So we minimise the sum of residuals Sum of U-hat$_i$ = SUM(Y$_i$ – Y-hat$_i$) making it as small as possible

Econometrics

But this won't work e.g. +10, -10; +2, -2 = sum of errors = 0.

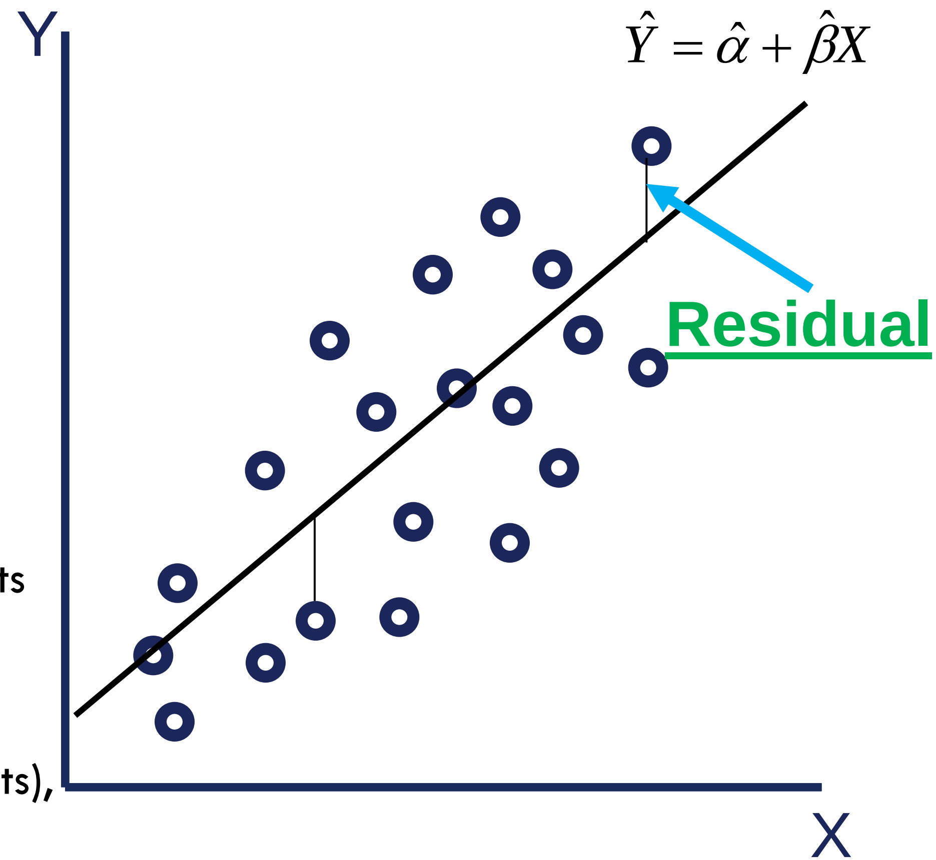
So we use the **LEAST SQUARES METHOD**

Sum of (u-hat)^2 = sum (y – yhat)^2
= Sum (y – α-hat – $\beta_1$hat-X)^2

By squaring, you give more weights to the bigger errors.

So now it is not possible for the Sum of the u-hats (the error terms) to be small even if u is widely spread.

OBVIOUSLY each time we change α and β_1 (hats), we will change the u-hat.
So we need to pick / find a α and β_1-hat such that the u-hat is as small as possible.



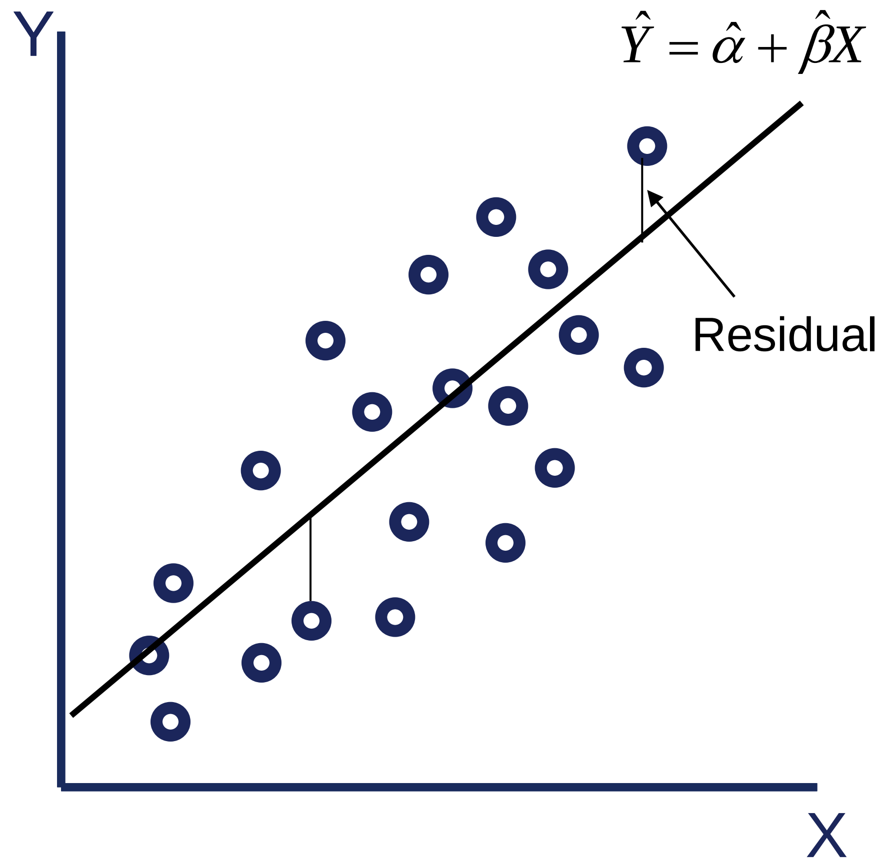
= OLS

OLS

$$\widehat{\beta}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

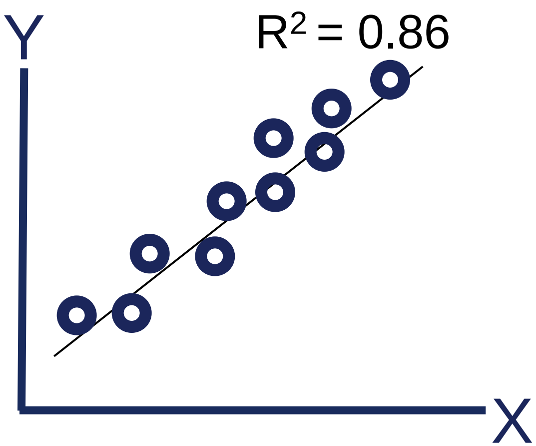
- When we fit a sample regression line to a scatter of points, it makes sense to select a line (that is, choose values for alpha-hat and beta-hat) such that the residuals given by that result are in some sense small.
- The most popular and best known way of ensuring this is to choose alpha-hat and beta-hat so as to minimise the sum of the squares of the residuals.
- This method of estimating the parameters alpha and beta is known as the method of ordinary least squares (OLS).
- Provided that a whole series of further assumptions are valid, the OLS method can be shown to provide ‘good’ estimators of alpha and beta.
- Computer software can compute OLS regressions automatically. The results may be reported as:

$$\hat{Y} = 30.71 + 0.812\, X$$

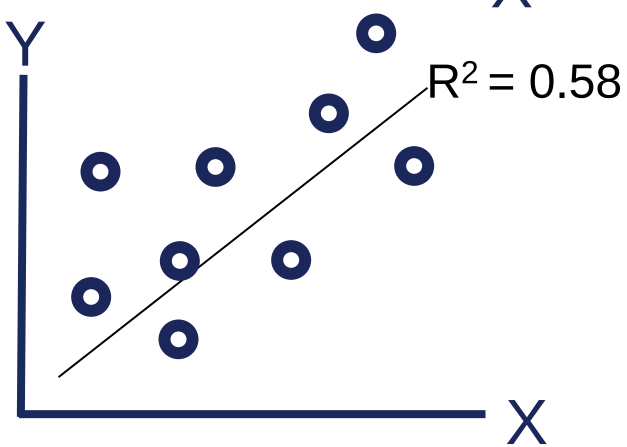


Measures of closeness of Fit

- The sample regression equation fits the scatter "fairly closely"
- But "fairly closely" is a vague expression, and it is often convenient to have a precise summary statistic (that is, a single number) by which we can assess and compare the closeness of fit of different scatters and different sample lines.



- We ask: What proportion of the variation in the consumption among our sample of households can we attribute to the variation in their incomes?
- We define the **coefficient of determination (R^2)** as the proportion of the sample variation in Y that can be attributed to the sample variation in X.
- The **closer the coefficient is to one**, the better the line will fit the points.



Testing for significance

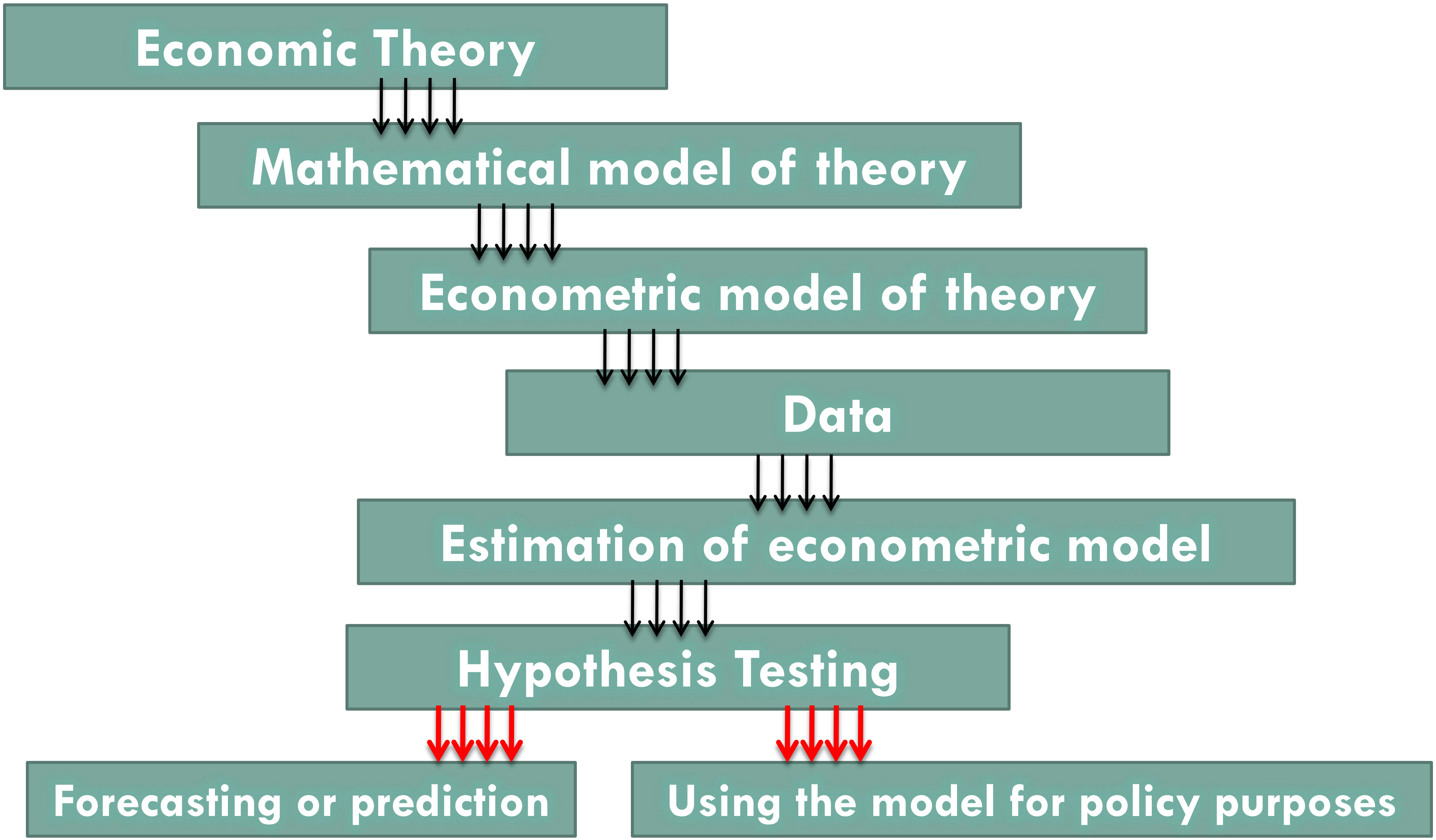
Does household disposable income have a **STATISTICALLY SIGNIFICANT** effect on household consumption?

- OLS regression will always try to fit a line through the points
- However we want to be able to test if our coefficient is significantly different from zero.
- To do this we calculate a t-value:

$$\mathbf{t = \beta/s.e.}$$

- That is, the t-value is the coefficient value divided by it's **standard error** (a measure of variance)
- This value can be looked up in statistical tables. However, as a rule of thumb a variable is significant if **the t-value is greater than 2**.

Modelling


Economic Theory
Mathematical model of theory
Econometric model of theory
Data
Estimation of econometric model
Hypothesis Testing
Forecasting or prediction
Using the model for policy purposes

Mis-specification…

- Once you have an economic theory, collected the data, you still have to come up with a mathematical model to test. Getting the mathematical model is CRUCIAL
- E.g. Data on Demand for Apples
- From microeconomic theory, it is known that the demand for a commodity generally depends on the real income of the consumer, the real price of the commodity, and the real prices of competing or complementary commodities.
- Specification of the model – what "shape" will this demand curve take? – what variables do I need to put into this specification?
  - E.g. real income; real price of apples; real price of oranges; real price of bananas
- Do I model it:
- Demand for Apples = $\alpha + \beta_1$ (Price of Apples) + β_2(Price of oranges) + β_3(Real income)
- Demand for Apples = $\alpha + \beta_1$Ln(Price of Apples) + β_2(Price of oranges+Price of bananas) + β_3(Real Income)^2

A word to the Econometrician…

- It should be clear that model building is an art as well as a science: The Ten Commandments of Applied Econometrics (Feldstein)

- 1) Thou shalt use common sense and economic theory
- 2) Thou shalt ask the right questions (i.e. put prevalence before mathematical elegance)
- 3) Thou shalt know the context (do not perform ignorant statistical analysis)
- 4) Thou shalt inspect the data
- 5) Thou shalt not worship complexity
- 6) Thou shalt look long and hard at the results
- 7) Thou shalt beware the costs of data mining
- 8) Thou shalt not confuse statistical significance with practical signifiance
- 9) Thou shalt use common sense and economic theory
- 10) Thou shalt not underestimate the task of data collection

AN INTRODUCTION TO ECONOMETRICS

Oxbridge Economics; Mo Tanweer
Mohammed.Tanweer@cantab.net